Edge-Constrained Joint View Triangulation for Image Interpolation

Maxime LHUILLIER Long QUAN
CNRS-GRAVIR-INRIA
655 avenue de l'Europe
38330 Montbonnot, France

Proc. IEEE CVPR'2000: vol.2, pages 218—224.

Abstract

Image-based-interpolation creates smooth and photo-realistic views between two view points. The concept of joint view triangulation (JVT) has been proven to be an efficient multi-view representation to handle visibility issue. However, the existing JVT, built only on a regular sampling grid, often produces undesirable artifacts for artificial objects. To tackle these problems, a new edge-constrained joint view triangulation is developed in this paper to integrate contour points and artificial rectilinear objects as triangulation constraints. Also a super-sampling technique is introduced to refine visible boundaries. The new algorithm is successfully demonstrated on many real image pairs.

1 Introduction

Image-based-interpolation [23, 2, 16, 11, 12] is gaining popularity for rendering smooth transition between different view points or for creating time-stopping effect [19]. The novel views directly rendered by images are more photo-realistic than classical CG geometric models [7, 22]. Reconstruction-based rendering methods including implicit reconstruction using matching tensors [1, 10] is more powerful as it may render images at arbitrary view point with very low computational cost, however reliable dense reconstruction is difficult to obtain with the actual state of art. In contrary, interpolation turns out more stable results [12] while operating only in image space.

As images are only samples of the actual 3D scenery instead of full representation of the scene, imaged based interpolation inherents the very difficult correspondence and occlusion problems, absent for computer generated images [2], [3]. This motivated the work [12] in which a five step algorithm was presented including an original joint view triangulation inspired by impostors [17] and mesh integration for range data [21, 18] to describe jointly the visibility of two views. The joint view triangulation in [12] gives only a rough structure which worked well for outdoor natural sceneries, but often have troubles with the manufactured objects including building, various kinds of posts ... This paper introduces a new edge-constrained joint view triangulation which integrate line segments from the polygonal approximation of contour points and an explicit modeling for artificial rectilinear objects. Another new feature is that the boundaries of the visible areas are refined by introducing a super-sampling grid.

The paper is organized as follows. In Section 2, we first review the quasi-dense matching algorithm. Then, the new edge-constrained joint view triangulation is described in Section 3. After the constrained joint view triangulation, we demonstrate this data representation on interpolating in-between images from two reference images in Section 4. Finally, some concluding remarks and future directions are given in Section 5.

2 Review of quasi-dense disparity map construction

In different images, matching either high-level image primitives such as feature points and line segments or just pixels is probably the hardest practical problem for vision applications. It has been particularly studied for a stereo rig in which the relative orientation reduces the search space from the 2D image plane to 1D along epipolar lines [9, 5]. Meanwhile the state of the art on matching does not yet give very satisfied general results. In fact, almost all matching algorithms have troubles either on occlusion or untextured areas. This is not surprising as there is just no enough information available in these areas which allow to make decision. This has motivated the definition of a quasi-dense matching [12]. The key remark is that the disparity map could never be everywhere dense, the best we can hope is only a set of sparsely distributed dense regions. This quasi-dense disparity map such defined is therefore a more realistic goal.

The construction of quasi-dense disparity map starts from matching some points of interest which have the highest textureness as seed points to bootstrap a region growing type algorithm to propagate the matches in its neighborhood from the most textured (therefore most reliable) pixels to less textured ones. The algorithm could therefore be described in two steps: Seed selection and propagation.

Seed selection Points of interest [15, 13, 8] are naturally good seed point candidates, as points of interest are by its very definition image points which have the highest textureness, i.e. the local maxima of the auto-correlation function of the signal.

So we first extract points of interest [15] from two original images, then a ZNCC correlation method is used to match the points of interest across two images, followed by a cross validation for the pair of images. This gives the initial list of correspondences sorted by the correlation score.

Propagation From the current seed list which is initialized by the first step, at each step, we pull the best match from the list of seeds. Then we look for additional matches in the neighborhood of the best seed. The neighbors of a seed point is taken to be all pixels within the 5×5 window centered at it. For each neighboring pixel of the first image, we first construct in the second image a list of tentative match candidates which consists of all pixels of a 3×3 window in the neighborhood of its corresponding location in the second image (see Figure 1).

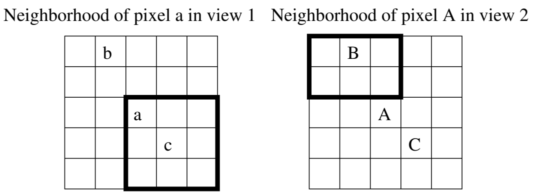

Figure 1. Definition of neighborhood $\mathcal{N}(a, A)$ of pixel match (a, A). It is a set of matches included in the two 5×5-neighborhood $\mathcal{N}_5(a)$ and $\mathcal{N}_5(A)$ of pixels a and A. Possible matches for b (resp. C) are in the 3×3 black frame centered at B (resp. c). The complete definition of $\mathcal{N}(a, A)$ is $\{(b, B), b \in \mathcal{N}_5(a), B \in \mathcal{N}_5(A), (B - A) - (b - a) \in \{-1, 0, 1\}^2\}$.

A such match neighborhood enforces the continuity constraint and a disparity gradient limit of one pixel for the matching result. The matching criterion is still the ZNCC correlation score but within a 5×5 window, therefore ghosting artifacts at the occluding contours are limited.

Finally additional matches in the neighborhood of the current seed pair are added simultaneously in the match list and the seed list such that the unicity constraint is preserved. The algorithm terminates when the seed match list becomes empty.

This algorithm could be efficiently implemented with a heap data structure for the seed pixels of the regions of the matched points. Notice that as each time only the best match is selected, this drastically limits the possibility of bad matches. For instance, the seed selection step seems very similar to many existing methods [24, 20] for matching points of interest using correlation, but the crucial difference is that we need only to take the most reliable ones rather than trying to match a maximum of them. In some extreme cases, only one good match of points of interest is sufficient to provoke an avalanche of the whole textured images. This makes our algorithm much less vulnerable. The same is true for propagation, the risk of bad propagation is considerable diminished by the best first strategy over all matched boundary points.

3 Edge-Constrained Joint View Triangulation

Image interpolation relies exclusively on image content with no depth information, triangulation is always necessary for computation efficiency and for that the matching is hardly anywhere dense. However, classical independent triangulation, operated in each individual image, lacks content consistency between different images. Inspired by jointly triangulating a set of range data in [18], the idea of triangulating simultaneously and consistently two images was first introduced for image interpolation in [12] as joint view triangulation to handle the visibility. The algorithm proposed in [12] is built merely on a regular sampling grid. It simplifies the implementation, but gives poor approximation for the visual events. Natural outdoor scenes could be nicely handled, but the algorithm often produces undesirable artifacts for artificial objects. The most typical artifact is the '*broken line*' phenomenon illustrated in Figure 2. To tackle these problems, a new edge-constrained joint view triangulation is developed in this section by introducing edge constraints and by explicitly modeling artificial rectilinear objects. Also a super-sampling technique is also introduced to refine the boundaries of visible areas.

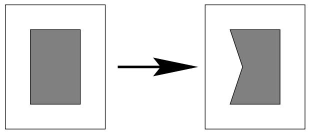

Figure 2. The 'broken line' artifact: The original image of a gray box on a white background is broken on the border due to non-appropriate mesh.

The consistency for the edge-constrained joint view triangulation is defined as follows.

1. There is one-to-one correspondence between the matched image points;

2. There is one-to-one constrained Delaunay edge correspondence in two images. There are three types of Delaunay edge constraints:

   - **Boundary edges of matched areas** The matched areas from the quasi-dense disparity map represent the common visible visual events of the scenes. Their boundary should not be crossed by any other triangles, they should be preserved and are natural constraints [12].
   - **Line segments** The most direct remedial measures for broken line artifact is to integrate image contour points as constraints. The line segments are polygonal approximation of linked contour points.
   - **Artificial rectilinear objects** Artificial rectilinear objects such as electric posts, trunks, ... are very frequent in outdoor scenes. They are often mishandled due to the size of the sampling grid which is limited by stability consideration of fitting. When their detection and matching is possible, they are also integrated into Delaunay constraints. An explicit modeling of these objects is presented in the following section.

3. The triangulation in each image is a constrained Delaunay triangulation due to its minimal roughness property [14].

3.1 Algorithm

Putting all these constraints together, the new edge-constrained joint view triangulation is implemented as an incremental insertion algorithm in both images simultaneously, consisting of five major steps. Figure 3 illustrates the evolution of the construction algorithm.

1. **Initialization**

   The initial empty images are triangulated by the diagonal from the top left to bottom right corner as illustrated in Figure 3.

2. **Detection of visible patches**

   The raw quasi-dense disparity map could not be used directly as visible pixels as it may still be corrupted, irregular and not structured. The goal is to use local geometric constraints encoded by a planar transformation to create reliable visible patches as visible content.

   The first image plane is initially divided into a regular grid of 8×8 pixels squares. The scale is a trade-off between the sampling resolution and regularisation stability. To increase the resolution a super-sampling grid by shifting the first grid half-size of the grid width is also introduced as illustrated in Figure 4.

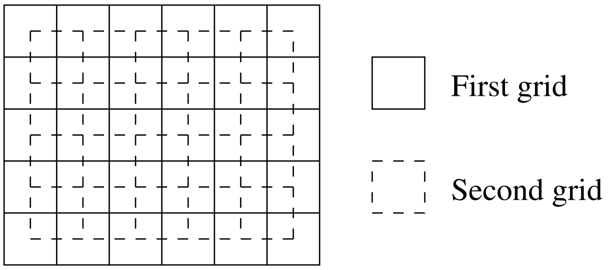

Figure 4. Two overlapping regular grids to subdivide the first image into square patches.

   For each square patch, all matched points inside it from the disparity map are obtained. A plane transformation is tentatively fitted to these matched points. The most general linear plane transformation is a homography represented by a homogeneous 3×3 non singular matrix. Four matched points, no three of them collinear, are sufficient to estimate a plane homography. Notice that an affine transformation encoded by 6 d.o.f. rather than a homography could also be used if the perspective distortion is moderate between images.

   Because a textured patch is rarely a perfect planar facet except for manufactured objects, the putative transformation for a patch can not be estimated by standard least squares estimators. Robust methods have to be adopted, which provide a reliable estimate of the homography even if some of the matched points of the square patch are not actually lying on the common plane on which the majority lies. The Random Sample Consensus (RANSAC) method originally introduced by Fischler and Bolles [6] is used for robust estimation of the homography. RANSAC has been successfully used for robust computation of the geometric matching tensors in [20, 24]. If the consensus for the transformation reaches 75%, the square patch is considered as a valid planar patch. The delimitation of the corresponding planar patch in the second image is defined by mapping the four corners of the grid in the first image with the estimated homography. This process of fitting the square patch to a homography is first repeated for all square patches of the first image for the two sampling grids.

   Each valid planar patch will be divided into 2 trian-

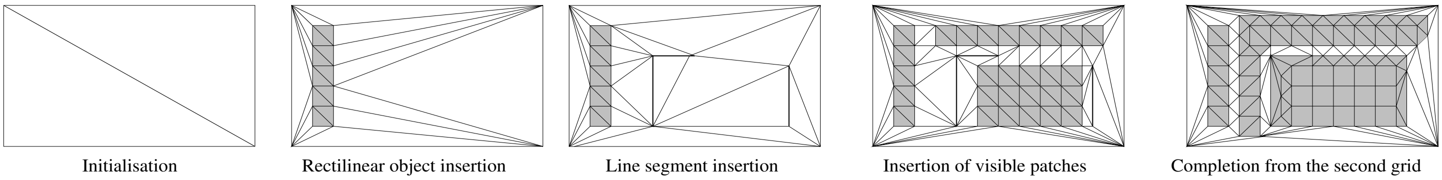

Figure 3. Illustration of the different steps of the edge-constraint joint view triangulation.

gles along one of its diagonals for further processing.

3. **Detection and insertion of rectilinear structures**

   The vertical rectilinear structures are modeled as sets of connected visible patches. First all vertical triplets of patches are detected, then all overlapping triplets are merged into connecting groups. Finally each group is completed by the adjacent and vertical square containing a line segment. The whole procedure is illustrated in Figure 5.

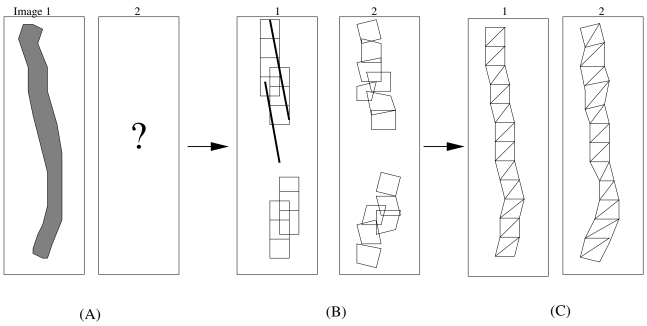

Figure 5. Illustration of detection of a rectilinear object (**left**) whose width is about the size of actual sampling grid. **Middle**: all corresponding vertical triplets of visible patches without left and right patch neighbors are first detected. **Right**: all overlapping triplets of patches form a connecting group. The vertical squares which share a line segment with the group are attached to the group.

   These groups of connecting patches are inserted into the current joint triangulation. A real example for detecting a small wood post on the foreground and the consequent triangulation is shown in Figure 6.

4. **Detection and insertion of line segments**

   The contour points as local maxima of gradient edge points are first detected by a Canny-like detector [4], then connecting contour points are linked and approximated by line segments. This procedure is only performed on the first image. We do not wish to directly match line segments in two images as such a procedure is hardly stable. A deduction method illustrated in Figure 7 is implemented which deduces the corre-

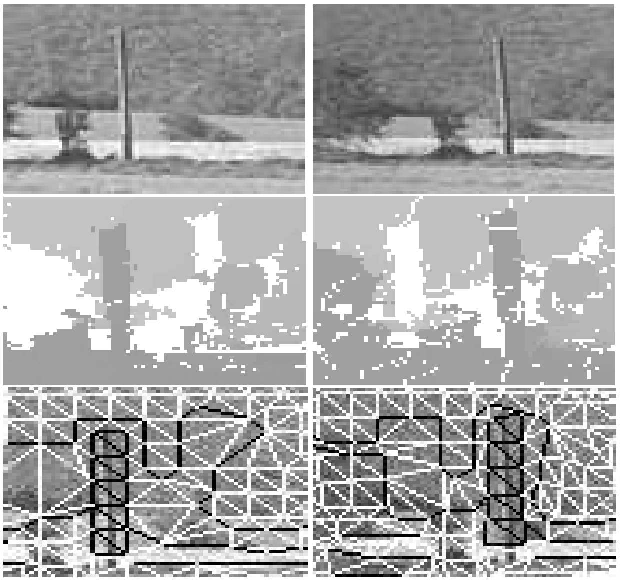

Figure 6. **Top**: two views of a small wood post. **Middle**: The disparity map after propagation with epipolar constraint used. **Bottom**: The modeling of the small post and the final JVT in which constraint edges are drawn in black.

sponding line segment in the second image from the disparity map.

For any line segment detected in the first image, all corresponding points in the neighborhood of the endpoints of the line segment are obtained from the disparity map. A tentative 1D transformation on the line segment in two images is estimated. This 1D transformation might be projective encoded by a 2×2 homography by taking two endpoints and one midpoint obtained from the epipolar geometry. If the epipolar geometry is not available, an affine transformation defined by two endpoints is estimated.

The second stage of the algorithm re-samples the line segment by regularly dividing it into smaller parts. Each segment part is validated by taking into account the number of the match points around it from the disparity map. The final line segment is selected as the one which maximizes the number of matched segment parts.

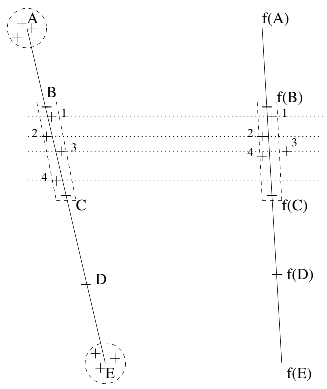

Figure 7. Illustrative figure for the deduction of corresponding line segments. AE is a line segment detected in the first image. A tentative corresponding line segment f(A)f(E) is defined by the neighboring points of A and E so that a 1D homography f is fitted to all contour points along the segment. The segment AE is regularly divided into smaller parts AB,BC,CD,DE and each one is validated by counting the number of matches which satisfy f in their neighborhood. For instance, matches 1 and 2 satisfy the homography while 3 and 4 do not. The segment f(A)f(E) which maximizes the number of accepted smaller parts is accepted to define the validated parts, for instance BC and f(B)f(C).

The corresponding line segment is inserted in the current joint triangulation while not violating the existing constraints. If it crosses any existing constraint edge, the line segment is further splitted for inserting only its non-intersecting parts.

5. **Insertion of visible patches from first sampling grid**

   All visible patches from the first sampling grid which do not intersect with any existing constraint edges are inserted into the current joint view triangulation.

6. **Boundary refinement from the super-sampling grid**

   First, all vertices of visible patches from the shifted grid are inserted if they are outside the current visible patches in the joint view triangulation. The triangles touching the current boundaries are considered as visible ones whose edges are the new boundaries of the matched areas if the match propagation succeeds in these triangles.

4 Examples

The new edge-constraint joint view triangulation algorithm has been demonstrated on many real image pairs. Mpeg sequences of interpolated images could be played at our Web site ***. All in-between images parameterized by $\lambda \in [0, 1]$ is generated using the same pseudo-painter's algorithm as [12].

Figure 8. Two frames of the original garden flower sequence.

Figure 9 shows a comparative result on two images from the public domain *flower garden* mpeg sequence (Figure 8). We can easily notice three broken line artifacts: one on the upper part of the tree, one on the middle of the tree and one on the background oblique white post. Most of these artifacts are removed with the edge-constrained method except the one in the middle of the tree persists.

One typical example of a pair of outdoor building images is illustrated in Figure 10. This example is difficult for several reasons.

- The leaves are hardly matched as the texture disparity is big and there is also shortage of seed matches.
- The epipolar lines are almost oriented the same way as the scene horizontal direction. It makes the matching of vertical structure easier while it does not help matching the horizontal structures of the building.

The matching error for horizontal edges of the building and pavement may reach several pixels.

- The upper sky area and the lower road area are almost textureless.

Very good results are obtained as shown in Figure 10. We can notice that in addition to the two front posts, two 'false posts' are also created on the background on the right side of the image, but it makes no damage for the interpolation results illustrated in Figure 11.

Figure 9. **Top**: joint view triangulation without edge constraint of the *flower garden* image pair. **Middle**: joint view triangulation with edge constraint. **Bottom**: one frame of the interpolated sequence with (right) and without (left) edge constraint. The edges in black are the constraints and those in white are Delaunay. For JVT without edge constraint, only the borders of the matched areas are constrained edges, broken line artifacts appear on the upper and middle part of the front tree and in the middle of the white oblique post on the background. For JVT with edge constraints, the above mentioned artifacts are considerably reduced, only on the middle part of the front tree persists.

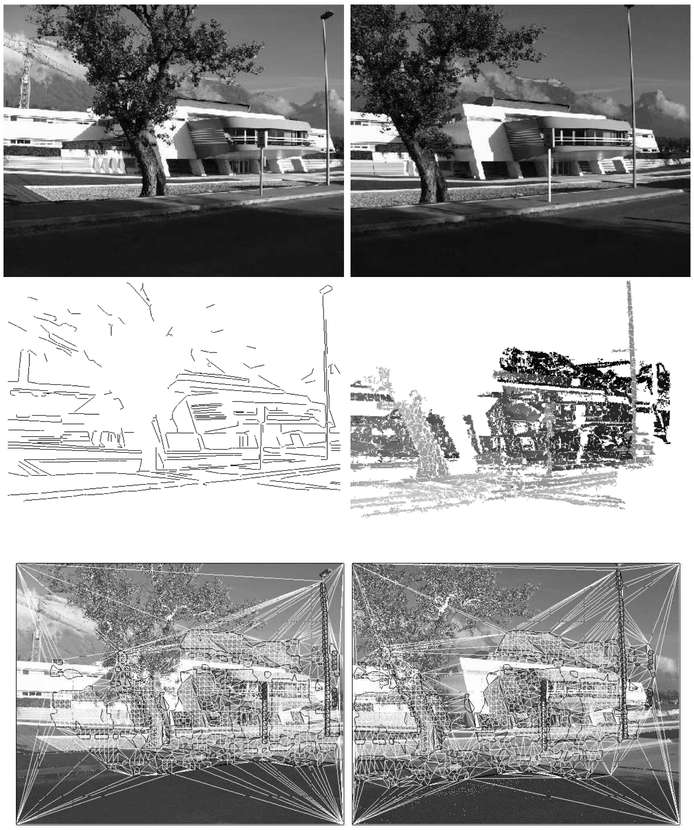

Figure 10. **Top**: the original image pair. **Middle**: the left is the edge image of the first frame and the right one is the disparity map. **Bottom**: the constraint joint view triangulation in two images.

Due to space limitation, we show only the interpolation results of another example in Figure 12.

5 Conclusion and future work

We have presented a new edge-constrained joint view triangulation for interpolating two images. Unlike the existing joint view triangulation which builds only on a regular sampling grid, the new algorithm integrates line segments and

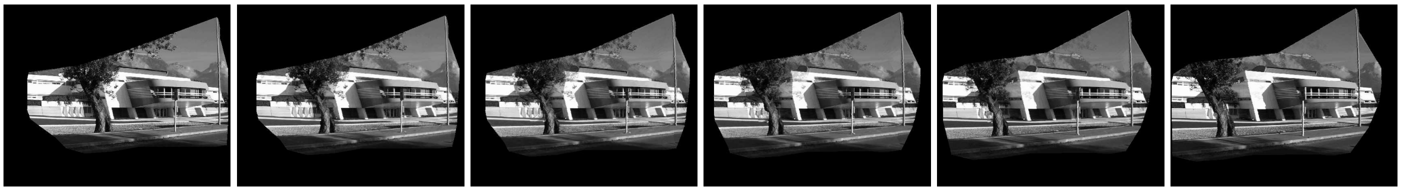

Figure 11. Some samples of the interpolated views at $\lambda = 0, 0.2, 0.4, 0.6, 0.8, 1$ from left to right. The Mpeg file is also available for reviewers as **Movie 1**.

Figure 12. Some samples of the interpolated views at $\lambda = 0.2, 0.4, 0.6, 0.8$ from left to right for a outdoor street scene. The Mpeg file is also available for reviewers as **Movie 2**.

explicit artificial rectilinear objects. A super-sampling grid is also used to refine the mesh boundaries for the visible areas. The quality of rendering results is improved in occluding contours where human eyes are extremely sensitive. We are actually investigate more advanced super-sampling methods to the refinement of the boundaries and working on multi-view joint view triangulation algorithms.

References

[1] S. Avidan and A. Shashua. Novel view synthesis in tensor space. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 1034–1040, June 1997.

[2] T. Beier and S. Neely. Feature-based image metamorphosis. In SIGGRAPH *1992*, pages 35–42, 1992.

[3] S.E. Chen and L. Williams. View interpolation for image synthesis. In SIGGRAPH *1993*, pages 279–288. Computer Graphics, 1993.

[4] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *International Journal of Computer Vision*, 1(2):167–187, 1987.

[5] U.R. Dhond and J.K. Aggarwal. Structure from stereo – a review. IEEE *Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, November 1989.

[6] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Graphics and Image Processing*, 24(6):381 – 395, June 1981.

[7] J.D. Foley, A. van Dam, S.K. Feiner, and J.F. Hughes. *Computer Graphics : Principles and Practice*. Addison-Wesley, November 1991.

[8] C. Harris and M. Stephens. A combined corner and edge detector. In *Alvey Vision Conference*, pages 147–151, 1988.

[9] A. Koschan. What is new in computational stereo since 1989 : A survey on stereo papers. Technical report, Department of Computer Science, University of Berlin, August 1993.

[10] S. Laveau and O. Faugeras. 3D scene representation as a collection of images and fundamental matrices. Technical report, INRIA, February 1994.

[11] S.Y. Lee, K.Y. Chwa, J. Hahn, and S.Y. Shin. Image morphing using deformation techniques. *The Journal of Visualization and Computer Animation*, 7:3–26, 1996.

[12] M. Lhuillier and L. Quan. Image interpolation by joint view triangulation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, volume 2, pages 139–145, 1999.

[13] B.D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence*, 1981.

[14] D. Rippa. Minimal roughness property of the delaunay triangulation. *Computer Aided Geometric Design*, 7:489–497, 1990.

[15] C. Schmid, R. Mohr, and Ch. Bauckhage. Comparing and evaluating interest points. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 230–235, January 1998.

[16] S.M. Seitz and C.R. Dyer. View morphing. In *Proceedings of* SIGGRAPH*, New Orleans, LA*, pages 21–30, 1996.

[17] F. Sillion, G. Drettakis, and B. Bodelet. Efficient impostor manipulation for real-time visualization of urban scnenery. In *Eurographics'97, Barcelona, Spain*, 1997.

[18] M. Soucy and D. Laurendeau. A general surface approach to the integration of a set of range views. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(4):344–358, april 1995.

[19] D. Taylor. Virtual camera movement: The way of the future ? *American Cinematographer Magazine Online*, 77(9):93–100, September 1996.

[20] P.H.S. Torr and A. Zisserman. Robust parameterization and computation of the trifocal tensor. In R.B. Fisher and E. Trucco, editors, *Proceedings of the seventh British Machine Vision Conference, Edinburgh, Scotland*, volume 2, pages 655–664. British Machine Vision Association, September 1996.

[21] G. Turk and M. Levoy. Zippered polygon meshes from range images. In SIGGRAPH *1994*, pages 311–318, 1994.

[22] A. Watt and M. Watt. *Advanced Animation and Rendering Techniques*. ACM Press, 1992.

[23] G. Wolberg. *Digital Image Warping*. IEEE Computer Society Press, Los Alamitos, California, 1990.

[24] Z. Zhang, R. Deriche, O.D. Faugeras, and Q.T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial Intelligence*, 78(1-2):87–119, 1994. Appeared in October 1995, also INRIA Research Report No.2273, May 1994.